

Mixing Time of Glauber Dynamics With Parallel Updates and Heterogeneous Fugacities

Mathieu Leconte, Jian Ni, and R. Srikant*

Abstract

Glauber dynamics is a powerful tool to generate randomized, approximate solutions to combinatorially difficult problems. Applications include Markov Chain Monte Carlo (MCMC) simulation and distributed scheduling for wireless networks. In this paper, we derive bounds on the mixing time of a generalization of Glauber dynamics where multiple vertices are allowed to update their states in parallel and the fugacity of each vertex can be different. The results can be used to obtain various conditions on the system parameters such as fugacities, vertex degrees and update probabilities, under which the mixing time grows polynomially in the number of vertices.

1 Introduction

Consider a graph $G = (V, E)$, where V is the set of vertices and E is the set of edges. Suppose $|V| = n$. For each vertex $v \in V$, we use $\mathcal{N}_v = \{w \in V : (v, w) \in E\}$ to denote the set of neighbors of v in the graph. An *independent set* of G is a subset of the vertices where no two vertices are neighbors of each other. Let \mathcal{I} be the set of all independent sets of G .

A *configuration* of the vertices in G is a vector of the form $(\sigma_v)_{v \in V}$, with $\sigma_v \in \Lambda = \{0, 1\}$ for all $v \in V$. For a vertex v and a configuration $\sigma \in \Lambda^n$, we say $v \in \sigma$ if $\sigma_v = 1$. A configuration σ on G is *feasible* if the set $\{v \in V : \sigma_v = 1\}$ is an independent set of G , i.e., if

$$\sigma_v + \sigma_w \leq 1, \text{ for all } (v, w) \in E. \quad (1)$$

Let $\Omega \subseteq \Lambda^n$ be the set of all feasible configurations on G .

We associate each vertex $v \in V$ with a parameter λ . We are interested in the following *product-form distribution* over the feasible configurations (independent sets) of the graph:

$$\pi(\sigma) = \frac{1}{Z} \prod_{v \in \sigma} \lambda, \quad (2)$$

$$Z = \sum_{\sigma \in \Omega} \prod_{v \in \sigma} \lambda. \quad (3)$$

Note that this corresponds to the so-called *hard-core gas model* studied in statistical physics, where λ is called the *fugacity* (e.g., [6, 9]).

*The authors are with the Coordinated Science Laboratory and Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA (leconte2@illinois.edu, jianni@illinois.edu, rsrikant@illinois.edu).

Glauber dynamics is a Markov chain which generates the stationary distribution in (2). It has many applications in statistical physics and computer science (e.g., hard-core gas model, graph coloring, approximate counting, combinatorial optimization [2, 3, 6]). Under (single-site) Glauber dynamics, in each time slot one vertex is selected uniformly at random, and only that vertex can change its state while other vertices keep their states unchanged. Let $\sigma(t)$ be the state of the Markov chain in time slot t .

Single-Site Glauber Dynamics (in Time Slot t)

1. Choose a vertex $v \in V$ uniformly at random.
 2. For vertex v :
 - If** $\sum_{w \in \mathcal{N}_v} \sigma_w(t-1) = 0$
 - (a) $\sigma_v(t) = 1$ with probability $p = \frac{\lambda}{1+\lambda}$.
 - (b) $\sigma_v(t) = 0$ with probability $\bar{p} = \frac{1}{1+\lambda}$.
 - Else**
 - (c) $\sigma_v(t) = 0$.
 3. For any vertex $w \in V \setminus \{v\}$:
 - (d) $\sigma_w(t) = \sigma_w(t-1)$.
-

It is not hard to verify that the Glauber dynamics Markov chain is *reversible* and has the product-form distribution in (2). In most applications, the performance of the Glauber dynamics is determined by how fast the Markov chain converges to the stationary distribution. The Glauber dynamics is said to have the *fast (rapid) mixing property* if the mixing time is polynomial in the size of the graph (the number of vertices n). In [9] it was shown that single-site Glauber dynamics has a mixing time of $O(n \log n)$ when $\lambda < \frac{2}{\Delta-2}$, where Δ is the maximum vertex degree in the graph.

Recently, Glauber dynamics has been applied to design distributed throughput-optimal scheduling algorithms for wireless networks (e.g., [5, 7, 8]). In the wireless network setting, the graph $G = (V, E)$ corresponds to the *interference graph* of the wireless network, where the vertices in V represent *links* (transmitter-receiver pairs) in the network, and there is an edge between two vertices in G if the corresponding wireless network links interfere with each other. A feasible schedule of the network is a set of links which do not interfere with each other, which corresponds to an independent set in the interference graph G . To achieve maximum throughput, the fugacities need to be chosen as appropriate functions of the queue lengths of the links, which are normally different from link to link. This motivates the study of Glauber dynamics with *heterogenous fugacities*.

In addition, in wireless networks, potentially multiple network links (vertices in G) can update their states in a single time slot, and we would expect that the mixing time of the Glauber dynamics Markov chain will be reduced with such parallel updates. However, the Markov chain may not even have the product-form distribution (which is a key property for establishing throughput-optimality in [5, 7, 8]) if we let an arbitrary set of vertices update their states. The following two questions then arise:

- (1) *How to select the vertices in each time slot to update their states such that the product-form distribution is maintained?*
- (2) *What is the mixing time of such a Glauber dynamics with parallel updates?*

The first question has been addressed in [7] and the second question will be addressed in this paper. The rest of the paper is organized as follows. In Section 2 we introduce a generalization of Glauber dynamics with parallel updates and heterogenous fugacities. In Section 3 we provide some technical background on the mixing time of Markov chains. In Sections 4 and 5 we derive bounds on the mixing time of the Glauber dynamics with parallel updates and heterogenous fugacities. The paper is concluded in Section 6.

2 Glauber dynamics with parallel updates

In [7] we have introduced a generalization of Glauber dynamics where multiple vertices (wireless network links) are allowed to update their states in a single time slot, under which the Markov chain is reversible and retains the product-form distribution. The key idea is that in every time slot, we select an independent set of vertices $\mathbf{m} \in \mathcal{I}$ to update their states according to a distributed randomized procedure, i.e., we select $\mathbf{m} \in \mathcal{I}$ with probability $q_{\mathbf{m}}$, where $\sum_{\mathbf{m} \in \mathcal{I}} q_{\mathbf{m}} = 1$. We call \mathbf{m} the *update set* (or *decision schedule* in [7]). The parallel Glauber dynamics is formally described as follows.

Parallel Glauber Dynamics (in Time Slot t)

1. Randomly choose an update set $\mathbf{m} \in \mathcal{I}$ with probability $q_{\mathbf{m}}$.
 2. For all vertex $v \in \mathbf{m}$:
 - If** $\sum_{w \in \mathcal{N}_v} \sigma_w(t-1) = 0$
 - (a) $\sigma_v(t) = 1$ with probability $p_v = \frac{\lambda_v}{1+\lambda_v}$.
 - (b) $\sigma_v(t) = 0$ with probability $\bar{p}_v = \frac{1}{1+\lambda_v}$.
 - Else**
 - (c) $\sigma_v(t) = 0$.
 - For all vertex $w \notin \mathbf{m}$:
 - (d) $\sigma_w(t) = \sigma_w(t-1)$.
-

The following results on the parallel Glauber dynamics have been established in [7].

Lemma 1 *Let $\mathbf{m}(t)$ be the update set selected in time slot t . If $\sigma(t-1) \in \Omega$ and $\mathbf{m}(t) \in \mathcal{I}$, then $\sigma(t) \in \Omega$.*

Because $\sigma(t)$ only depends on the previous state $\sigma(t-1)$ and some randomly selected update set $\mathbf{m}(t)$, $\sigma(t)$ evolves as a discrete-time Markov chain (DTMC). Next we will derive the transition probabilities between the states.

Lemma 2 *A state $\sigma \in \Omega$ can make a transition to a state $\eta \in \Omega$ if and only if $\sigma \cup \eta \in \Omega$ and there exists an update set $\mathbf{m} \in \mathcal{I}$ with $q_{\mathbf{m}} > 0$ such that*

$$\sigma \triangle \eta = (\sigma \setminus \eta) \cup (\eta \setminus \sigma) \subseteq \mathbf{m},$$

and in this case the transition probability from σ to η is given by:

$$P(\sigma, \eta) = \sum_{\mathbf{m} \in \mathcal{I}: \sigma \triangle \eta \subseteq \mathbf{m}} q_{\mathbf{m}} \left(\prod_{v \in \sigma \setminus \eta} \bar{p}_v \right) \left(\prod_{v \in \eta \setminus \sigma} p_v \right) \left(\prod_{v \in \mathbf{m} \cap (\sigma \cap \eta)} p_v \right) \left(\prod_{v \in \mathbf{m} \setminus (\sigma \cup \eta) \setminus \mathcal{N}_{\sigma \cup \eta}} \bar{p}_v \right). \quad (4)$$

Under the Glauber dynamics with parallel updates, let

$$q_v = \sum_{\mathbf{m} \ni v} q_{\mathbf{m}}$$

be the probability of updating vertex v in a time slot.

Theorem 1 *A necessary and sufficient condition for the Markov chain of the parallel Glauber dynamics to be irreducible and aperiodic is $\cup_{\mathbf{m} \in \mathcal{I}: q_{\mathbf{m}} > 0} \mathbf{m} = V$, or equivalently, $q_v > 0$ for all $v \in V$, and in this case the Markov chain is reversible and has the following product-form stationary distribution:*

$$\pi(\sigma) = \frac{1}{Z} \prod_{v \in \sigma} \lambda_v, \quad (5)$$

$$Z = \sum_{\sigma \in \Omega} \prod_{v \in \sigma} \lambda_v. \quad (6)$$

Remark 1 *The single-site Glauber dynamics can be viewed as a special case of the parallel Glauber dynamics in which $q_{\mathbf{m}} > 0$ if and only if the update set \mathbf{m} always consists of only one vertex.*

In this paper we will show that the parallel Glauber dynamics has a very fast mixing time $O(\log n)$ under certain conditions for bounded-degree graphs. On the other hand, it was shown in [4] that the single-site Glauber dynamics has a mixing time at least $\Omega(n \log n)$ for bounded-degree graphs.

3 Mixing Time of Markov Chains

Consider a finite-state, irreducible, aperiodic Markov chain (P, Ω, π) where P denotes the transition matrix, Ω denotes the state space, and π denotes the unique stationary distribution of the Markov chain.

Definition 1 *The variation distance between two distributions μ, ν on Ω is defined as*

$$\|\mu - \nu\|_{var} = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|. \quad (7)$$

Definition 2 *The mixing time $T_{mix}(\epsilon)$ for $\epsilon > 0$ of the Markov chain is defined as the time required for the Markov chain to get close to the stationary distribution. More precisely,*

$$T_{mix}(\epsilon) = \max_{x \in \Omega} \inf \left\{ t : \|P^t(x, \cdot) - \pi\|_{var} \leq \epsilon \right\}. \quad (8)$$

Definition 3 A coupling of the Markov chain is a stochastic process $(X(t), Y(t))$ on $\Omega \times \Omega$ such that $\{X(t)\}$ and $\{Y(t)\}$ marginally are copies the original Markov chain, and if $X(t) = Y(t)$, then $X(t+1) = Y(t+1)$.

Let Φ be a distance function (*metric*) defined on $\Omega \times \Omega$, which satisfies that for any $x, y, z \in \Omega$:

- (1) $\Phi(x, y) \geq 0$, with equality if and only if $x = y$.
- (2) $\Phi(x, y) = \Phi(y, x)$.
- (3) $\Phi(x, z) \leq \Phi(x, y) + \Phi(y, z)$.

Let

$$D_{min} = \min_{x, y \in \Omega, x \neq y} \Phi(x, y), \quad D_{max} = \max_{x, y \in \Omega} \Phi(x, y), \quad D = \frac{D_{max}}{D_{min}}.$$

The following result can be used to obtain an upper bound for the mixing time of the Markov chain (e.g., [2]).

Theorem 2 Suppose there exist a constant $\beta < 1$ and a coupling $(X(t), Y(t))$ of the Markov chain such that, for all $x, y \in \Omega$,

$$E[\Phi(X(t+1), Y(t+1)) | X(t) = x, Y(t) = y] \leq \beta \Phi(x, y). \quad (9)$$

Then the mixing time of the Markov chain is bounded by:

$$T_{mix}(\epsilon) \leq \frac{\log(D\epsilon^{-1})}{1 - \beta}. \quad (10)$$

In general, determining β is hard since one needs to check the contraction condition (9) for all pairs of configurations. In [1] the so-called *path coupling* method was introduced by Bubley and Dyer to simplify the calculation. Under path coupling, we only need to check the contraction condition for certain pairs of configurations. The path coupling method is described in the following theorem.

Theorem 3 Let $S \subseteq \Omega \times \Omega$ and suppose for all $X, Y \in \Omega \times \Omega$, there exists a path $X = Z_0, Z_1, \dots, Z_r = Y$ between X and Y such that $(Z_l, Z_{l+1}) \in S$ for $0 \leq l < r$ and

$$\Phi(X, Y) = \sum_{l=0}^{r-1} \Phi(Z_l, Z_{l+1}).$$

Suppose there exist a constant $\beta < 1$ and a coupling $(X(t), Y(t))$ of the Markov chain such that for any $(x, y) \in S$,

$$E[\Phi(X(t+1), Y(t+1)) | X(t) = x, Y(t) = y] \leq \beta \Phi(x, y).$$

Then the mixing time of the Markov chain is bounded by:

$$T_{mix}(\epsilon) \leq \frac{\log(D\epsilon^{-1})}{1 - \beta}. \quad (11)$$

Note that the key simplification in the path coupling theorem (Theorem 3), compared to the coupling theorem (Theorem 2), is that the contraction condition (9) needs to hold only for $(x, y) \in S$, instead of $(x, y) \in \Omega \times \Omega$.

4 Mixing Time of Glauber Dynamics With Parallel Updates

In this section we analyze the mixing time of Glauber dynamics with parallel updates using the path coupling theorem. We will use the following distance function: for any $\sigma, \eta \in \Omega$,

$$\Phi(\sigma, \eta) = \sum_v |\sigma_v - \eta_v| f(v) = \sum_{v \in \sigma \Delta \eta} f(v), \quad (12)$$

where $f(v) > 0$ is a (weight) function of $v \in V$ and recall that $\sigma \Delta \eta = (\sigma \setminus \eta) \cup (\eta \setminus \sigma)$. Note that this distance function is a weighted Hamming distance function and satisfies all the properties of a metric.

Consider the following *coupling* $(\sigma(t), \eta(t))$: in every time slot both chains select the same update set and use the same coin toss for every vertex in the update set if that vertex can be added to both configurations.

Let $E[\Delta\Phi(\sigma(t), \eta(t))]$ be the (conditional) expected change of the the distance between the states of the two Markov chains $\{\sigma(t)\}$ and $\{\eta(t)\}$ after one slot:

$$E[\Delta\Phi(\sigma(t), \eta(t))] = E[\Phi(\sigma(t+1), \eta(t+1)) | \sigma(t), \eta(t)] - \Phi(\sigma(t), \eta(t)).$$

For any $\mathbf{m} \in \mathcal{I}$, let

$$E[\Delta^{\mathbf{m}}\Phi(\sigma(t), \eta(t))] = E[\Delta\Phi(\sigma(t), \eta(t)) | \mathbf{m} \text{ is the update set}].$$

Lemma 3 *Let $\tilde{\mathbf{m}} = (y_1, \dots, y_{|\mathbf{m}|})$ be any ordering of \mathbf{m} . For any $\sigma(t), \eta(t) \in \Omega$,*

$$E[\Delta^{\mathbf{m}}\Phi(\sigma(t), \eta(t))] = \sum_{k=1}^{|\mathbf{m}|} E[\Delta^{y_k}\Phi(\sigma(t), \eta(t))].$$

Proof Note that the value of $\Phi(\sigma, \eta)$ is completely determined by the set $\sigma \Delta \eta$, which in turn depends only on σ and η . Hence, it suffices to show that we will obtain the same sets $\sigma(t+1)$ and $\eta(t+1)$ by updating all the vertices of \mathbf{m} simultaneously and by updating them in any sequential order.

The moves trying to remove vertices from a configuration will be successful in all cases. The outcome of a move trying to add a vertex y to a configuration ω is successful if and only if $\mathcal{N}_y \cap \omega = \emptyset$. But \mathbf{m} is an independent set, so no neighbor of a vertex $y \in \mathbf{m}$ is in \mathbf{m} . Then the states of the neighbors of y are unchanged after updating any subset of vertices of \mathbf{m} . Hence, $\forall k \in \{1, \dots, |\mathbf{m}|\}$, the outcome of a move trying to add y_k to ω will be the same if we update the vertices of \mathbf{m} sequentially and if we update all the vertices of \mathbf{m} simultaneously. This is in particular true for the configurations $\sigma(t)$ and $\eta(t)$, so we conclude that the two update procedures will yield the same sets $\sigma(t+1)$ and $\eta(t+1)$.

We say that $\sigma, \eta \in \Omega$ are *adjacent* and we write $\sigma \sim \eta$ if there exists $v \in V$ such that σ and η differ only at v . Let

$$S = \{(\sigma, \eta) : \sigma, \eta \in \Omega \text{ and } \sigma \sim \eta\}.$$

Note that under the distance function defined in (12), for all $\sigma, \eta \in \Omega$, we can find a path $\sigma = \tau_0, \tau_1, \dots, \tau_{|\sigma \Delta \eta|} = \eta$ between σ and η such that $(\tau_l, \tau_{l+1}) \in S$ for $0 \leq l < r$ and $\Phi(\sigma, \eta) = \sum_{l=0}^{r-1} \Phi(\tau_l, \tau_{l+1})$.

Now consider a pair of adjacent configurations $\sigma(t)$ and $\eta(t)$ that differ only at v . Without loss of generality, suppose $\sigma_v(t) = 0$ and $\eta_v(t) = 1$. This means that, $\eta_w(t) = 0$ for all $w \in \mathcal{N}_v$. Since, $\sigma(t)$ and $\eta(t)$ differ only at v , this also means that $\sigma_w(t) = 0$ for all $w \in \mathcal{N}_v$.

Lemma 4

$$E[\Delta\Phi(\sigma(t), \eta(t))] \leq -q_v f(v) + \sum_{w \in \mathcal{N}_v} \frac{q_w \lambda_w}{1 + \lambda_w} f(w). \quad (13)$$

Proof Using Lemma 3, we have

$$\begin{aligned} E[\Delta\Phi(\sigma(t), \eta(t))] &= E_{\mathbf{m}} \left[E[\Delta^{\mathbf{m}}\Phi(\sigma(t), \eta(t))] \right] \\ &= \sum_{\mathbf{m}} q_{\mathbf{m}} E[\Delta^{\mathbf{m}}\Phi(\sigma(t), \eta(t))] \\ &= \sum_{\mathbf{m}} q_{\mathbf{m}} \sum_{y \in \mathbf{m}} E[\Delta^y\Phi(\sigma(t), \eta(t))] \\ &= \sum_{y \in V} q_y E[\Delta^y\Phi(\sigma(t), \eta(t))]. \end{aligned}$$

Note that only updates on vertices v and $w \in \mathcal{N}_v$ can affect the value of $E[\Delta\Phi(\sigma(t), \eta(t))]$. In particular, if v is selected for update and since we use the same coin toss for both Markov chains, $\sigma(t+1) = \eta(t+1)$. Thus $E[\Delta^v\Phi(\sigma(t), \eta(t))] = -f(v)$.

If $w \in \mathcal{N}_v$ is selected for update, under configuration $\eta(t)$, w can only take value 0 because w has a neighbor (i.e., v) belongs to $\eta(t)$. While under configuration $\sigma(t)$, there are two cases:

- 1) if w has a neighbor in $\sigma(t)$, then w can only take value 0;
- 2) if w has no neighbors in $\sigma(t)$, w can take value 1 with probability $\frac{\lambda_w}{1+\lambda_w}$ and value 0 otherwise.

Hence for $w \in \mathcal{N}_v$,

$$E[\Delta^w\Phi(\sigma(t), \eta(t))] \leq \frac{\lambda_w}{1 + \lambda_w} f(w).$$

Summing up all contributions we have (13).

Now we are ready to present the main result of this paper.

Theorem 4 For any positive function $f(v)$ of $v \in V$, let $m = \min_{v \in V} f(v)$, $M = \max_{v \in V} f(v)$, and $\xi = \frac{M}{m}$. If

$$\theta \triangleq \min_{v \in V} \left\{ q_v f(v) - \sum_{w \in \mathcal{N}_v} \frac{q_w \lambda_w}{1 + \lambda_w} f(w) \right\} > 0, \quad (14)$$

then the mixing time of the parallel Glauber dynamics is bounded by:

$$T_{mix}(\epsilon) \leq \frac{M}{\theta} \log(\epsilon^{-1} n \xi). \quad (15)$$

Proof For any pair of adjacent configurations $(\sigma(t), \eta(t)) \in S$ that differ at some vertex $v \in V$, from (13) and (14) we have:

$$E[\Delta\Phi(\sigma(t), \eta(t))] \leq -\theta \leq -\frac{\theta}{M} \Phi(\sigma(t), \eta(t)),$$

where we use the fact that

$$\Phi(\sigma(t), \eta(t)) = f(v) \leq M.$$

Therefore,

$$E[\Phi(\sigma(t+1), \eta(t+1)) | \sigma(t), \eta(t)] \leq \left(1 - \frac{\theta}{M}\right) \Phi(\sigma(t), \eta(t)).$$

Then, by applying the path coupling theorem where we let $\beta = 1 - \frac{\theta}{M}$ and $D = n\xi$, we prove the bound in (15).

4.1 Conditions for Fast Mixing of Parallel Glauber Dynamics

We can now specify the (weight) function f to obtain different conditions on the fugacities λ_v 's for fast mixing. We will show three such examples.

Corollary 1 *Let $m = \min_{v \in V} \frac{1+\lambda_v}{q_v}$, $M = \max_{v \in V} \frac{1+\lambda_v}{q_v}$, and $\xi = \frac{M}{m}$. If*

$$\theta \triangleq \min_{v \in V} \left\{ 1 + \lambda_v - \sum_{w \in \mathcal{N}_v} \lambda_w \right\} > 0, \quad (16)$$

then we have

$$T_{mix}(\epsilon) \leq \frac{M}{\theta} \log(\epsilon^{-1} n \xi). \quad (17)$$

Proof Choose $f(v) = \frac{1+\lambda_v}{q_v}$, $\forall v \in V$.

Corollary 2 *Let $q_{min} = \min_{v \in V} q_v$, $q_{max} = \max_{v \in V} q_v$, and $\xi = \frac{q_{max}}{q_{min}}$. If*

$$b \triangleq \max_{v \in V} \sum_{w \in \mathcal{N}_v} \frac{\lambda_w}{1 + \lambda_w} < 1, \quad (18)$$

then we have

$$T_{mix}(\epsilon) \leq \frac{\log(\epsilon^{-1} n \xi)}{q_{min}(1-b)}. \quad (19)$$

Proof Choose $f(v) = \frac{1}{q_v}$, $\forall v \in V$.

Remark 2 *If $q_v > c$ for some constant $c > 0$ which is independent of the size of the network n , then the mixing time is $O(\log n)$. In particular, for a bounded-degree graph G where δ and Δ are the minimum and maximum vertex degrees, if $q_v = \frac{1}{d_v+1}$ where d_v is the degree of v , then we have $\frac{1}{\Delta+1} \leq q_v \leq \frac{1}{\delta+1}$ and*

$$T_{mix}(\epsilon) \leq \frac{\Delta+1}{1-b} \log\left(\frac{\Delta+1}{\delta+1} \epsilon^{-1} n\right).$$

Corollary 3 *If $\lambda_v < \frac{1}{d_v-1}$ for all $v \in V$, then $T_{mix}(\epsilon) \leq \frac{M}{\theta} \log(\epsilon^{-1} n \xi)$, for some constants M , ξ and $\theta > 0$.*

Proof The proof is very similar to that of Theorem 4. We can choose $f(v) = \frac{d_v}{q_v}$ and let $M = \max_{v \in V} \frac{d_v}{q_v}$ and

$$\xi = \frac{\max_{v \in V} \frac{d_v}{q_v}}{\min_{v \in V} \frac{d_v}{q_v}}.$$

The parallel Glauber dynamics will have fast mixing if

$$\theta = \min_{v \in V} \left\{ d_v - \sum_{w \in \mathcal{N}_v} \frac{\lambda_w}{1 + \lambda_w} d_w \right\} > 0.$$

To achieve that we need, for all $v \in V$,

$$d_v - \sum_{w \in \mathcal{N}_v} \frac{\lambda_w}{1 + \lambda_w} d_w > 0.$$

It is sufficient that $\frac{\lambda_v}{1 + \lambda_v} d_v < 1$, which is equivalent to $\lambda_v < \frac{1}{d_v - 1}$.

Note that the condition $\lambda_v < \frac{1}{d_v - 1}$ for all $v \in V$ might be very different from $b = \max_{v \in V} \sum_{w \in \mathcal{N}_v} \frac{\lambda_w}{1 + \lambda_w} < 1$ (e.g., in a star network).

5 Mixing Time of Single-Site Glauber Dynamics With Heterogeneous Fugacities

Since the single-site Glauber dynamics is a special case of the parallel Glauber dynamics (in which the update set always consists of only one vertex), the general results derived in the previous section also apply to the single-site Glauber dynamics. The motivation of this section is to derive a larger region on the fugacities under which the single-site Glauber dynamics is fast mixing, where we use a similar path coupling technique as in [9].

We redefine the state space to be the set of all configurations Λ^n . For a vertex $v \in V$ and a configuration $\sigma \in \Lambda^n$, we define the set of *blocked* neighbors of v with respect to σ as

$$B_\sigma(v) = \left\{ w \in \mathcal{N}_v : w \in \sigma \text{ or } \mathcal{N}_w \cap \sigma \neq \emptyset \right\}$$

and of *unblocked* neighbors of v as

$$\bar{B}_\sigma(v) = \mathcal{N}_v \setminus B_\sigma(v).$$

Note that the unblocked neighbors of v are the neighbors of v that can be added to the configuration in the next move. If $v \notin \sigma$, we write σ^v to denote the configuration that differs from σ only at v .

We say that $\sigma, \eta \in \Lambda^n$ are *adjacent* and we write $\sigma \sim \eta$ if there exists $v \in V$ such that σ and η differ only at v . Let

$$S = \left\{ (\sigma, \eta) : \sigma \sim \eta \right\}.$$

Then, for all $\sigma, \eta \in \Lambda^n$, we can define a path in Λ^n between σ and η as a sequence $(\tau_0, \dots, \tau_r) \subseteq \Lambda^n$ such that for $0 \leq i < r$, $\tau_i \sim \tau_{i+1}$ and $\tau_0 = \sigma, \tau_r = \eta$. Let $\mathcal{P}(\sigma, \eta)$ be the set of all paths in Λ^n between σ and η .

We will use the following distance function on $\Lambda^n \times \Lambda^n$: $\forall \sigma, \eta \in \Lambda^n$, let

$$\Phi(\sigma, \eta) = \min_{(\tau_0, \dots, \tau_r) \in \mathcal{P}(\sigma, \eta)} \sum_{i=0}^{r-1} l(\tau_i, \tau_{i+1}),$$

where

$$l(\sigma, \sigma^v) = 1 + \frac{1}{2} \sum_{w \in \bar{B}_\sigma(v)} \lambda_w \quad (20)$$

can be viewed as the *length* of edge (σ, σ^v) according to the adjacent relationship defined by S . Φ is clearly symmetric, non-negative, zero only when the configurations are identical, and satisfies the triangle inequality (because $\mathcal{P}(\sigma, \eta) \subseteq \mathcal{P}(\sigma, \mu) \times \mathcal{P}(\mu, \eta)$). Hence, it is indeed a metric on $\Lambda^n \times \Lambda^n$.

For all $v, u \in V$, we let $\mathcal{T}(v, u) = \mathcal{N}_v \cap \mathcal{N}_u$ be the set of vertices that form triangles with v and u . To each vertex $v \in V$, we associate a fugacity λ_v .

Consider a pair of adjacent configurations $\sigma \sim \sigma^v$ (where $v \notin \sigma$). Suppose $\sigma(t) = \sigma$, $\eta(t) = \sigma^v$. The *coupling* is simply that each configuration attempts the same move at every time slot. More precisely, both Markov chains select the same vertex to update, and use the same coin toss if the vertex can be added to the configurations. For convenience, we use the following notations from [9]. Let

$$E[\Delta\Phi] = E[\Phi(\sigma(t+1), \eta(t+1)|\sigma(t), \eta(t)) - \Phi(\sigma(t), \eta(t))].$$

This can be further calculated via the analysis of individual moves. Let

$$\begin{aligned} E[\Delta^{+y}\Phi] &= E[\Delta\Phi | \text{both chains attempt to add } y \text{ at time } t], \\ E[\Delta^{-y}\Phi] &= E[\Delta\Phi | \text{both chains attempt to remove } y \text{ at time } t], \end{aligned}$$

and denote the total effect of all moves on y by

$$E[\Delta^y\Phi] = \frac{\lambda_y}{1 + \lambda_y} E[\Delta^{+y}\Phi] + \frac{1}{1 + \lambda_y} E[\Delta^{-y}\Phi].$$

We provide the main result of this section in the following theorem, which says that the mixing time of (single-site) Glauber dynamics with heterogenous fugacities is $O(n \log n)$ under certain conditions.

Theorem 5 *Let*

$$a = \max_{v \in V} \left\{ \sum_{w \in \mathcal{N}_v} \lambda_w \right\} \quad (21)$$

and $\gamma = \sum_{y \in V} (1 + \lambda_y)$. If $a < 2$ and the probability of selecting vertex $y \in V$ to update is $q_y = \frac{1 + \lambda_y}{\gamma}$, then the mixing time of the Glauber dynamics is bounded by

$$T_{\text{mix}}(\epsilon) \leq \frac{\gamma}{1 - \frac{a}{2}} \log \left(\epsilon^{-1} n \left(1 + \frac{a}{2} \right) \right). \quad (22)$$

We need the following lemma to prove Theorem 5, and its proof is given in the Appendix.

Lemma 5

$$2 \sum_{y \in V} (1 + \lambda_y) E[\Delta^y\Phi] \leq -2 + \sum_{w \in \mathcal{N}_v} \lambda_w. \quad (23)$$

Proof (Theorem 5) Suppose $a < 2$ and $q_y = \frac{1+\lambda_y}{\gamma}$, then using Lemma 5 we have

$$\begin{aligned}
E[\Delta\Phi] &= \sum_{y \in V} q_y E[\Delta^y \Phi] \\
&= \frac{\sum_{y \in V} (1 + \lambda_y) E[\Delta^y \Phi]}{\gamma} \\
&\leq \frac{-1 + \frac{a}{2}}{\gamma} \\
&\leq \frac{-1 + \frac{a}{2}}{\gamma} \Phi(\sigma(t), \eta(t)),
\end{aligned}$$

where we use the fact that $\Phi(\sigma(t), \eta(t)) \geq 1$ for any $\sigma(t) \neq \eta(t)$. Therefore,

$$E[\Phi(\sigma(t+1), \eta(t+1)) | \sigma(t), \eta(t)] \leq (1 + \frac{-1 + \frac{a}{2}}{\gamma}) \Phi(\sigma(t), \eta(t)).$$

Since $\Phi(\sigma, \sigma^v) \geq 1$, so for all $\sigma, \eta \in \Lambda^n$, $\sigma \neq \eta$, we have $\Phi(\sigma, \eta) \geq 1$. Also, $\Phi(\sigma, \sigma^v) \leq 1 + \frac{a}{2}$, and thus $\Phi(\sigma, \eta) \leq n(1 + \frac{a}{2})$. Then, apply the path coupling theorem where we let $\beta = 1 + \frac{-1 + \frac{a}{2}}{\gamma}$ and $D = n(1 + \frac{a}{2})$:

$$\begin{aligned}
T_{\text{mix}}(\epsilon) &\leq \frac{1}{1 - (1 + \frac{-1 + \frac{a}{2}}{\gamma})} \log \left(n(1 + \frac{a}{2}) \epsilon^{-1} \right) \\
&= \frac{\gamma}{1 - \frac{a}{2}} \log \left(\epsilon^{-1} n(1 + \frac{a}{2}) \right) \\
&\leq \frac{3n}{1 - \frac{a}{2}} \log(2\epsilon^{-1}n).
\end{aligned}$$

Example 1 Consider a graph G with n vertices, labelled $1, 2, \dots, n$. For each vertex i , let $\mathcal{N}_i = \{j : j \neq i \text{ and } |j - i| \leq 3\}$ be the set of neighbors of i . For Glauber dynamics with homogeneous fugacity λ , Vigoda's result [9] says that it has mixing time $O(n \log n)$ when $\lambda < \frac{2}{\Delta-2} = 0.5$, since the maximum vertex degree $\Delta = 6$ in this graph. While for Glauber dynamics with heterogeneous fugacities, our result says that even the fugacity at some vertex exceeds 0.5 (but is less than 2), the Glauber dynamics can still have $O(n \log n)$ mixing time, e.g., when $\lambda_i = 1$ for vertices $i = 6l + 1$, $l = 0, 1, \dots$, and $\lambda_j = 0.18$ for other vertices j .

6 Conclusion

In this paper we have analyzed the mixing time of a generalization of Glauber dynamics with parallel updates and heterogeneous fugacities. By applying the path coupling theorem and by choosing appropriate distance functions, we have obtained various conditions on the system parameters such as fugacities, vertex degrees and update probabilities under which the parallel Glauber dynamics is fast mixing. In particular, we have shown that the mixing time of the parallel Glauber dynamics grows as $O(\log n)$ in the number of vertices n for bounded-degree graphs when the fugacities satisfy certain conditions.

A Proof of Lemma 5

Proof Note that when $a < 2$, the edge length defined in (20) satisfies $1 \leq l(\sigma, \sigma^v) < 2$. Hence the length of a path of two edges or more is at least 2. This implies that the distance

between two adjacent configurations is simply the length of the direct edge between them, i.e.,

$$\Phi(\sigma, \sigma^v) = l(\sigma, \sigma^v) = 1 + \frac{1}{2} \sum_{w \in \bar{B}_\sigma(v)} \lambda_w.$$

Only updates on vertices within a distance of 2 hops from v can influence the value of $\Phi(\sigma, \sigma^v)$. Hence,

$$\begin{aligned} & 2 \sum_{y \in V} (1 + \lambda_y) E[\Delta^y \Phi] \\ = & 2(1 + \lambda_v) E[\Delta^v \Phi] + 2 \sum_{w \in \mathcal{N}_v} (1 + \lambda_w) E[\Delta^w \Phi] + 2 \sum_{x \in \mathcal{N}_w : w \in \mathcal{N}_v} (1 + \lambda_x) E[\Delta^x \Phi]. \end{aligned}$$

We will examine separately the terms involving v , w 's and x 's. It is important to notice that the moves trying to remove a vertex from a configuration are always successful, whereas the moves trying to add a vertex y to a configuration σ succeed only if $\mathcal{N}_y \cap \sigma = \emptyset$.

$E[\Delta^v \Phi]$:

Only if $\mathcal{N}_v \cap \sigma = \emptyset$, will the move trying to add v to σ be successful. Therefore,

$$2(1 + \lambda_v) E[\Delta^v \Phi] = \begin{cases} -(1 + \lambda_v)(2 + \sum_{w \in \bar{B}_\sigma(v)} \lambda_w) & \text{if } \mathcal{N}_v \cap \sigma = \emptyset, \\ -(2 + \sum_{w \in \bar{B}_\sigma(v)} \lambda_w) & \text{otherwise.} \end{cases}$$

$E[\Delta^w \Phi]$:

Trying to update a neighbor of v can only increase the distance between the two configurations. Indeed, either adding a neighbor of v will fail, or it will increase the length of any path between the two configurations by one, because it is then necessary to remove that neighbor from both configurations before it may be possible to add v . Removing a neighbor of v from the configuration will potentially unblock some other neighbors of v and thus increase the distance between the configurations.

More precisely: if $w \in B_\sigma(v)$, it is impossible to add w to σ and so $E[\Delta^{+w} \Phi] = 0$; otherwise, $E[\Delta^{+w} \Phi] = \Phi(\sigma_w, \sigma^v) - \Phi(\sigma, \sigma^v)$. By the triangle inequality,

$$\Phi(\sigma_w, \sigma^v) \leq \Phi(\sigma_w, \sigma_{w,v}) + \Phi(\sigma_{w,v}, \sigma^v).$$

We have $\Phi(\sigma_w, \sigma_{w,v}) \leq \Phi(\sigma, \sigma^v)$ because $\bar{B}_{\sigma_w}(v) \subseteq \bar{B}_\sigma(v)$. Also, $2\Phi(\sigma_{w,v}, \sigma^v) = 2 + \sum_{y \in \bar{B}_{\sigma_{w,v}}(w)} \lambda_y$ and $\bar{B}_{\sigma^v}(w) = \bar{B}_\sigma(w) \setminus \{v\} \setminus \mathcal{T}(v, w)$. Then, by combining everything, we get

$$2E[\Delta^{+w} \Phi] \leq \begin{cases} 0 & \text{if } w \in B_\sigma(v), \\ 2 + \sum_{y \in \bar{B}_\sigma(w) \setminus \{v\} \setminus \mathcal{T}(v, w)} \lambda_y & \text{otherwise.} \end{cases}$$

Trying to remove w will succeed and unblock some of v 's neighbors, as well as add w to $\bar{B}_\sigma(v)$ if $\mathcal{N}_w \cap \sigma = \emptyset$. A neighbor w' of v will be unblocked if and only if w is the only neighbor of w' in σ . Thus,

$$2E[\Delta^{-w} \Phi] = \begin{cases} 0 & \text{if } w \notin \sigma, \\ \delta(\mathcal{N}_w \cap \sigma = \emptyset) \lambda_w + \sum_{w' \in \mathcal{T}(v, w) : \mathcal{N}_{w'} \cap \sigma = \{w\}} \lambda_{w'} & \text{otherwise.} \end{cases}$$

where we use the indicator function $\delta(a)$ which is equal to 1 if a is true and 0 otherwise.

$\mathbf{E}[\Delta^{\mathbf{x}}\Phi]$:

As in the case of the terms involving w 's, trying to remove a vertex x from the configurations unblocks some neighbors of v and thus increases the distance between the configurations. On the other hand, adding a vertex x blocks some neighbors of v and reduces the distance between the configurations. More precisely, when removing x from σ and σ^v , a neighbor w of v will be unblocked if and only if x is the only neighbor of w in σ . Thus,

$$2E[\Delta^{-x}\Phi] = \begin{cases} 0 & \text{if } w \notin \sigma, \\ \sum_{w \in \mathcal{T}(v,x): \mathcal{N}_w \cap \sigma = \{x\}} \lambda_w & \text{otherwise.} \end{cases}$$

Trying to add x fails if $x \in \sigma$ or $\mathcal{N}_x \cap \sigma \neq \emptyset$. Otherwise, it succeeds and blocks some neighbors of v . More precisely, a neighbor w of v is blocked if it is not in σ , it was not already blocked in configuration σ before the move, and it is a neighbor of x . Thus,

$$2E[\Delta^{+x}\Phi] = \begin{cases} 0 & \text{if } x \in \sigma \text{ or } \mathcal{N}_x \cap \sigma \neq \emptyset, \\ -\sum_{w \in \bar{B}_\sigma(v): x \in \mathcal{N}_w} \lambda_w & \text{otherwise.} \end{cases}$$

We can rewrite this equation in the following form:

$$2E[\Delta^{+x}\Phi] = - \sum_{w \in \bar{B}_\sigma(v): x \in \bar{B}_\sigma(w)} \lambda_w.$$

We are now ready to sum up the contributions from all the updates:

$$\begin{aligned} & 2 \sum_{y \in V} (1 + \lambda_y) E[\Delta^y \Phi] \\ & \leq -(1 + \lambda_v \delta(\mathcal{N}_v \cap \sigma = \emptyset)) \left(2 + \sum_{w \in \bar{B}_\sigma(v)} \lambda_w \right) \\ & + \sum_{w \in \bar{B}_\sigma(v)} \lambda_w \left(2 + \sum_{y \in \bar{B}_\sigma(w) \setminus \{v\} \setminus \mathcal{T}(v,w)} \lambda_y \right) \end{aligned} \tag{24}$$

$$+ \sum_{w \in \sigma} \left(\delta(\mathcal{N}_w \cap \sigma = \emptyset) \lambda_w + \sum_{w' \in \mathcal{T}(v,w): \mathcal{N}_{w'} \cap \sigma = \{w\}} \lambda_{w'} \right) \tag{25}$$

$$+ \sum_{x \in \sigma} \left(\sum_{w \in \mathcal{T}(v,x): \mathcal{N}_w \cap \sigma = \{x\}} \lambda_w \right) \tag{26}$$

$$- \sum_{x: d(v,x)=2} \lambda_x \left(\sum_{w \in \bar{B}_\sigma(v): x \in \bar{B}_\sigma(w)} \lambda_w \right). \tag{27}$$

We want to simplify the terms (24) + (27):

$$(24) + (27) = \sum_{w \in \bar{B}_\sigma(v)} \lambda_w \left(2 + \sum_{y \in \bar{B}_\sigma(w) \setminus \{v\} \setminus \mathcal{T}(v,w)} \lambda_y - \sum_{x \in \bar{B}_\sigma(w)} \lambda_x \right).$$

As we recall that the notation x in $\{x \in \bar{B}_\sigma(w)\}$ refers to vertices that are at distance 2 from v (as opposed to y that refers to any vertex in V), we get that

$$\{x \in \bar{B}_\sigma(w)\} = \bar{B}_\sigma(w) \setminus \{v\} \setminus \mathcal{T}(v, w),$$

that is the set $\bar{B}_\sigma(w)$ minus its elements at a distance 0 or 1 from v . Then, we can simplify:

$$(24) + (27) = 2 \sum_{w \in \bar{B}_\sigma(v)} \lambda_w.$$

We now turn to (25) + (26):

$$\begin{aligned} & (25) + (26) \\ &= \sum_{w \in \sigma} \left(\delta(\mathcal{N}_w \cap \sigma = \emptyset) \lambda_w + \sum_{w' \in \mathcal{T}(v, w): \mathcal{N}_{w'} \cap \sigma = \{w\}} \lambda_{w'} \right) + \sum_{x \in \sigma} \left(\sum_{w \in \mathcal{T}(v, x): \mathcal{N}_w \cap \sigma = \{x\}} \lambda_w \right) \\ &= \sum_{w \in \sigma: \mathcal{N}_w \cap \sigma = \emptyset} \lambda_w + \sum_{w \in B_\sigma(v): |\mathcal{N}_w \cap \sigma| = 1} \lambda_w \leq \sum_{w \in B_\sigma(v)} \lambda_w. \end{aligned}$$

Combining all these inequalities, we have

$$\begin{aligned} & 2 \sum_{y \in V} (1 + \lambda_y) E[\Delta^y \Phi] \\ &\leq -(1 + \lambda_v \delta(\mathcal{N}_v \cap \sigma = \emptyset)) \left(2 + \sum_{w \in \bar{B}_\sigma(v)} \lambda_w \right) + 2 \sum_{w \in B_\sigma(v)} \lambda_w + \sum_{w \in B_\sigma(v) \cup \sigma} \lambda_w \\ &= -2 + \sum_{w \in \mathcal{N}_v} \lambda_w - \lambda_v \delta(\mathcal{N}_v \cap \sigma = \emptyset) \left(2 + \sum_{w \in \bar{B}_\sigma(v)} \lambda_w \right) \\ &\leq -2 + \sum_{w \in \mathcal{N}_v} \lambda_w. \end{aligned}$$

References

- [1] R. Bubley and M. E. Dyer. Path coupling: A technique for proving rapid mixing in Markov chains. In *38th Annual Symposium on Foundations of Computer Science*, pages 223–231, October 1997.
- [2] M. Dyer and C. Greenhill. A more rapidly mixing Markov chain for graph colorings. *Random Structures & Algorithms*, 13(3-4):285–317, 1998.
- [3] M. Dyer and C. Greenhill. On Markov chains for independent sets. *Journal of Algorithms*, 35:17–49, 2000.
- [4] T. P. Hayes and A. Sinclair. A general lower bound for mixing of single-site dynamics on graphs. In *FOCS '05: Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, 2005.
- [5] L. Jiang and J. Walrand. A distributed CSMA algorithm for throughput and utility maximization in wireless networks. In *Proceedings 46th Annual Allerton Conference on Communication, Control and Computing*, September 2008.

- [6] F. Martinelli. Lectures on Glauber dynamics for discrete spin models. Lectures on probability theory and statistics (Saint-Flour, 1997), Lecture Notes in Math., 1717, Springer, Berlin, 1999.
- [7] J. Ni and R. Srikant. Distributed CSMA/CA algorithms for achieving maximum throughput in wireless networks. In *Information Theory and Applications Workshop*, February 2009. <http://arxiv.org/abs/0901.2333>.
- [8] S. Rajagopalan, D. Shah, and J. Shin. Network adiabatic theorem: an efficient randomized protocol for contention resolution. In *Proceedings of ACM Sigmetrics*, June 2009.
- [9] E. Vigoda. A note on the Glauber dynamics for sampling independent sets. *The Electronic Journal of Combinatorics*, 8, 2001.